

Compering the ACER and POT MCMC  
(acronyms?) Extreme Value Statistics  
Methods Through Analysis of Commodities  
Data

Kristoffer Kofoed Rødvei

April 2016

MASTER THESIS

Department of Production and Quality Engineering (What here??)

Norwegian University of Science and Technology

Supervisor: Professor Arivd Næss

Secondary Advisor: Professor Andre Riebler

Helping(???): Professor Sjur Westgaard

## Preface

Here, you give a brief introduction to your work. What it is (e.g., a Master's thesis in RAMS at NTNU as part of the study program xxx and...), when it was carried out (e.g., during the autumn semester of 2021). If the project has been carried out for a company, you should mention this and also describe the cooperation with the company. You may also describe how the idea to the project was brought up.

It is suggested that the reader have some statistical knowledge, especially within extreme value and computational statistical. Some financial knowledge and experience with value at risk is desirable.

"When your done: <https://www.scribbr.com/thesis/preface-thesis/>"

Trondheim, 2016-04-25

(Your signature)

Kristoffer Kofoed Rødvei

## **Summary and Conclusions**

Here you give a summary of your work and your results. This is like a management summary and should be written in a clear and easy language, without many difficult terms and without abbreviations. Everything you present here must be treated in more detail in the main report. You should not give any references to the report in the summary – just explain what you have done and what you have found out. The Summary and Conclusions should be no more than two pages.

You may assume that you have got three minutes to present to the Rector of NTNU what you have done and what you have found out as part of your thesis. (He is an intelligent person, but does not know much about your field of expertise.)

# Contents

Preface . . . . .	i
Summary and Conclusions . . . . .	ii
<b>1 Theory</b>	<b>2</b>
1.1 Extreme value theory . . . . .	2
1.1.1 Peak Over Threshold . . . . .	4
1.2 Average Condition Exceedance Rate . . . . .	5
1.3 Bayesian Inference . . . . .	5
1.4 Markov Chain Monte Carlo Method . . . . .	8
1.4.1 Gibbs Sampling . . . . .	8
1.4.2 Metropolis–Hastings Algorithm . . . . .	9
1.4.3 Effective Sample Size . . . . .	10
1.4.4 Adaptive Metropolis Algorithm . . . . .	11
1.5 Applying Markov Chain Monte Carlo to Peak Over Threshold . . . . .	12
1.5.1 Multivariate Random Normal Generator . . . . .	13
1.6 Value at Risk . . . . .	13
1.7 ARCH/GARCH . . . . .	13
<b>2 Data</b>	<b>14</b>
2.1 Synthetic Data . . . . .	14
2.1.1 Pareto . . . . .	14
2.1.2 Commodity Generated Data . . . . .	14
2.2 Commodity Data . . . . .	14

<i>CONTENTS</i>	1
<b>3 Summary</b>	<b>15</b>
3.1 Summary and Conclusions . . . . .	15
3.2 Discussion . . . . .	15
3.3 Recommendations for Further Work . . . . .	16
<b>Bibliography</b>	<b>17</b>

# Chapter 1

## Theory

This chapter gives an introduction of/to the theory behind this thesis/paper. Some of the following sections only give a brief description of the theory with sources for more in depth explanation. The main focus of this chapter is describing the theory behind the Markov chain Monte Carlo method for solving extreme value problem developed for this paper.

### 1.1 Extreme value theory

The basics of extreme value theory is to analyze

$$M_n = \max\{X_1, \dots, X_n\}. \quad (1.1)$$

It is assumed that  $X_1, \dots, X_n$  are independent and identically distributed (i.i.d) random variables with common cumulative distribution function  $F$ . The exact distribution of  $M_n$  then follows

$$\begin{aligned} \Pr(M_n \leq z) &= \Pr(X_1 \leq z, \dots, X_n \leq z) \\ &= \Pr(X_1 \leq z) \times \dots \times \Pr(X_n \leq z) \\ &= [F(z)]^n. \end{aligned} \quad (1.2)$$

The distributing  $F$  is normally unknown, and a divergent in the estimated distributing  $F$  could potentially escalate to a large difference in the resulting  $F^n$ .

As for the central limit theory for normal distribution,  $F^n$  also has a limiting distribution, by

The Fisher–Tippett–Gnedenko theorem. If there exist an  $a_n > 0$  and  $b_n$  such that

$$\lim_{n \rightarrow \infty} \Pr[(M_n - b_n)/a_n \leq z] \rightarrow G(z), \quad (1.3)$$

where  $G$  is a non-degenerating function, then  $G$  follows either

$$G(z) = \exp \left\{ -\exp \left[ -\left( \frac{z-b}{a} \right) \right] \right\}, \quad -\infty < z < \infty; \quad (1.4)$$

$$G(z) = \begin{cases} 0, & z \leq b, \\ \exp \left\{ -\left( \frac{z-b}{a} \right)^{-\alpha} \right\}, & z > b; \end{cases} \quad (1.5)$$

$$G(z) = \begin{cases} \exp \left\{ -\left[ -\left( \frac{z-b}{a} \right) \right]^\alpha \right\}, & z \leq b, \\ 1, & z > b; \end{cases} \quad (1.6)$$

for each case,  $\alpha > 0$ . Here equation (1.4), (1.5) and (1.6) refers to Gumbel, Fréchet and Weibull distribution respectively. The above equation can be combined into the General Extreme Value (GEV) distribution

$$G(z) = \exp \left\{ -\left[ 1 + \xi \left( \frac{z-\mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \quad (1.7)$$

where  $1 + \xi(z - \mu)/\sigma > 0$ , the location parameter is  $-\infty < \mu < \infty$ , the scale parameter is  $\sigma > 0$  and the shape parameter is  $-\infty < \xi < \infty$ . It can easily be verified that the GEV equals (1.5) when  $\xi > 0$ , (1.6) when  $\xi < 0$  and converge towards (1.4) as  $\xi \rightarrow 0$ .

The GEV model requires i.i.d data points for parameter estimations. Unfortunately, in practice this is rarely the case. Block maxima or  $r$  largest order statistics are common methods for converging the dependent data points into an approximate i.i.d dataset (Coles, 2001, Chapter 3). The basic principle is to only use the largest, or  $r$  largest data within each block. An example of a block could be week, month, year etc.

For a deeper description of extreme value theory, GEV or block maxima, the book of Coles (2001) is suggested.

### 1.1.1 Peak Over Threshold

One of the problem with the GEV method and the filtering of i.i.d data points, is that the block maxima and  $r$  largest order statistics are quite wasteful. Specially in situations where some blocks contains more extreme than others, large extreme will be discarded from the set, which else would have been accepted in other blocks.

The Peak Over Threshold (POT) method is suggesting a different method of tackling the i.i.d issue. Filtering the data, by only using points over a certain threshold  $u$ , avoid the problem of discarding large extremes in certain blocks. As long as  $u$  is chosen sufficiently large, the resulting data will be i.i.d. Using the GEV distribution it can be shown, like was done by (Coles, 2001, p. 76), that

$$\Pr(X > y + u | X > u) = \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \quad (1.8)$$

Which result in the distribution of  $y$

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}} \quad (1.9)$$

or

$$H(y) = 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right) \quad (1.10)$$

when  $\xi \rightarrow 0$ . Here  $\xi$  equal the GEV model, while  $\tilde{\sigma} = \sigma + \xi(u - \mu)$ . Since the requirement  $\tilde{\sigma} > 0$  still holds, for simplicity, the notation  $\sigma$  is used from here.

The conditional probability

$$\Pr(X > y + u | X > u) = \frac{\Pr(X > y + u)}{\Pr(X > u)}, \quad (1.11)$$

since  $y + u \geq u$ , the probability  $\Pr(X > y + u, X > u) = \Pr(X > y + u)$ . By combining equation (1.8) and (1.11), the probability of a future event can be found by

$$\begin{aligned} \Pr(X > z) &= \Pr(X > y + u) \\ &= \Pr(X > u) \cdot P(X > y + u | X > u) \\ &= \Pr(X > u) \cdot \left[1 + \xi \left(\frac{z - u}{\sigma}\right)\right]^{-\frac{1}{\xi}}, \end{aligned} \quad (1.12)$$



where  $\Pr(X > u)$  is the probability that a random point exceeds the threshold.

## Threshold

HOW TO FIND THRESHOLD

## 1.2 Average Condition Exceedance Rate

MORE blabla

## 1.3 Bayesian Inference

For the traditional frequentist statistics, the parameters  $\theta$  are assumed fixed, while data  $x_1, \dots, x_n$  are random from the underlying distribution  $f(\mathbf{x}|\theta)$ . Bayesian statistics, instead treats the parameters  $\theta$  with a probability distribution, where it is possible to make subjective beliefs about the distribution, independent of the data. These subjective beliefs is used to construct a prior distribution  $f(\theta)$  based on experiences, information or physical knowledge of the situation analyzed.

The posterior distribution of the parameters dependent on the measured data becomes

$$f(\theta|\mathbf{x}) = \frac{f(\theta)f(\mathbf{x}|\theta)}{\int_{\Theta} f(\theta)f(\mathbf{x}|\theta)d\theta}, \quad (1.13)$$

where  $\Theta$  is the domain over all possible parameters, for which the integral is taken, and  $f(\mathbf{x}|\theta)$  is the likelihood function. The likelihood function is constructed from the joint density function, which for independent data equals

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta). \quad (1.14)$$

The integral over parameters reduces to a constant, which makes

$$f(\theta|\mathbf{x}) \sim c \cdot f(\theta)f(\mathbf{x}|\theta), \quad (1.15)$$

where  $c = 1 / \int_{\Theta} f(\boldsymbol{\theta}) f(\mathbf{x}|\boldsymbol{\theta}) d\boldsymbol{\theta}$  is the normalizing constant.

A conjugate prior is a prior which combined with the likelihood function construct a posterior distribution in the same family as the prior. Conjugate priors are often preferred because of the analytical luxury and computational simplicity.

Since Bayesian inference accounts for the distribution of parameter, equation (1.12) can be rewritten using  $\Pr(X > u) = \alpha$

$$\Pr(X > z) = \alpha \left[ 1 + \xi \left( \frac{z - u}{\sigma} \right) \right]^{-\frac{1}{\xi}}, \quad (1.16)$$

where  $\alpha$ ,  $\xi$  and  $\sigma$  are the unknown parameters. Since  $\alpha$  is the probability of a point being larger than the threshold,  $\alpha$  is independent of  $\xi$  and  $\sigma$ . For the independent parameters, development of the posterior distributions can be treated separately.

Starting with  $\xi$  and  $\sigma$ , by combining equation (1.14) and (1.9), the joint density function for the POT method becomes

$$\begin{aligned} f(\mathbf{y}|\xi, \sigma) &= \prod_{i=1}^k h(y_i|\xi, \sigma) \\ &= \sigma^{-k} \prod_{i=1}^k \left( 1 + \frac{\xi y_i}{\sigma} \right)^{-\left(1 + \frac{1}{\xi}\right)}, \end{aligned} \quad (1.17)$$

or by (1.10),  $f(\mathbf{y}|\sigma) = \sigma^{-k} \exp \left\{ -\frac{1}{\sigma} \sum_{i=1}^k y_i \right\}$  when  $\xi = 0$ .

The obvious beginning for investigating priors is the conjugate priors, but unfortunately, there do not appear to be any conjugate priors for the joint POT probability distribution. This paper will not go into depth on how to select Bayesian priors for the POT method, but instead use the suggestion proposed by (Coles, 2001, p. 174). Note that there are potential improvements by deeper investigation of POT/GEV priors, especially for priors developed for specific situations where there are physical knowledge or practical experiences about the parameters. Since  $\sigma > 0$ , the transformation  $\phi = \log(\sigma)$  ensures  $\sigma$  to be valid without restriction on  $\phi$ . The suggested priors are  $f_{\phi}(\cdot)$  and  $f_{\xi}(\cdot)$  to be normally distributed around zero with variance  $\nu_{\phi} = 10^4$  and  $\nu_{\xi} = 100$ .

Considering the prior distribution of  $\phi$  instead of  $\sigma$ , the change of variable for the joint den-

sity function becomes

$$\begin{aligned}
 f_{\mathbf{y}|\xi,\phi}(\mathbf{y}|\xi, \phi) &= \frac{f_{\mathbf{y},\xi,\phi}(\mathbf{y}, \xi, \phi)}{f_{\xi,\phi}(\xi, \phi)} \\
 &= \frac{f_{\mathbf{y},\xi,\sigma}(\mathbf{y}, \xi, \exp(\phi)) \cdot \left| \frac{d}{d\phi} \exp(\phi) \right|}{f_{\xi,\sigma}(\xi, \exp(\phi)) \cdot \left| \frac{d}{d\phi} \exp(\phi) \right|} \\
 &= f_{\mathbf{y}|\xi,\sigma}(\mathbf{y}|\xi, \exp(\phi)).
 \end{aligned} \tag{1.18}$$

The posterior distribution of the parameters can then be developed by the priors, (1.18) and (1.15)

$$f_{\xi,\phi|\mathbf{y}}(\xi, \phi|\mathbf{y}) \sim c \cdot f_{\mathbf{y}|\xi,\sigma}(\mathbf{y}|\xi, \exp(\phi)) f_{\xi}(\xi) f_{\phi}(\phi), \tag{1.19}$$

where again  $c$  is the normalizing constant,  $f_{\mathbf{y}|\xi,\sigma}$  is as in (1.17),  $f_{\xi}(\xi) \sim N(0, 100)$  and  $f_{\phi}(\phi) \sim N(0, 10^4)$ .

The development of  $\alpha$  posterior distribution, can be started with investigating priors. Since  $\alpha$  simply equals  $\Pr(X > u)$ , the range is limited between 0 and 1. For simplicity the prior is set to  $f(\alpha) \sim UNIF(0, 1)$ . It is noted that the distribution of  $\alpha$  in reality is not flat. Low valued  $\alpha$  is more probably than high, while the probability converges to zero for the endpoints. A well-tuned Beta distributed prior could account for this, and improve the result.

The joint density function can be created by the fact that  $\alpha$  simply equals the probability of a random point exceeding the threshold. This can be expressed using the binominal distribution, using  $k_i$  for the numbers of points exceeding the threshold and  $N_i$  for the total numbers of point, each for a given period  $i$ . For a total  $n$  numbers of periods, the posterior distribution equals

$$\begin{aligned}
 f(\alpha|k_1, \dots, k_n, N_1, \dots, N_n) &\sim f(\alpha) \cdot \prod_{i=1}^n f(k_i|N_i, \alpha) \\
 &= \prod_{i=1}^n \binom{N_i}{k_i} \alpha^{k_i} (1 - \alpha)^{N_i - k_i} \\
 &\sim \alpha^{\sum_{i=1}^n k_i} (1 - \alpha)^{\sum_{i=1}^n N_i - \sum_{i=1}^n k_i}.
 \end{aligned} \tag{1.20}$$

The resulting distribution is independent of period selection, the notation  $k$  and  $N$  can be used

for the total numbers of exceedance and the total numbers of measurements respectively. It is noted that the distribution is proportional to the Beta distribution. Since there only exist one normalizing constant which fulfills the requirements for a probability distribution, the posterior distribution is not only proportional, but equal to the Beta distribution. Rewriting equation (1.20) gives

$$f(\alpha|k, N) \sim BETA(k + 1, N - k + 1). \quad (1.21)$$

## 1.4 Markov Chain Monte Carlo Method

Markov Chain Monte Carlo (MCMC) is a powerful iterative method used to sample from probability distributions, which analytically or through other simulation methods, could be difficult and impracticable to sample from. The algorithm is constructed by converging the desired probability distribution to an irreducible and aperiodic Markov Chain with limiting distribution equal the target distribution. Independent of starting position the Markov Chain will then converge towards the desired probability distribution in the limit as the numbers of iterations goes to infinity. The first numbers of realizations from the Markov Chain until converging is called the burn-in period, and is discarded for further analyses. The remaining realizations, approximately follows the desired probability distribution, where the accuracy increase as the numbers of realizations increase. Monte Carlo method can then be used to calculate the quantity of interest like mean, expected value, future prediction, credible interval etc. (Givens and Hoeting, 2013, p. 201)

(Gamerman and Lopes (2006)

(Givens and Hoeting, 2013, p.220)

### 1.4.1 Gibbs Sampling

In situations where it is difficult to sample from the joint distribution, but applicable from the conditional distribution, Gibbs sampler is preferable. The principle of Gibbs sampling is to construct the Markov Chain, by repeatedly sample each parameter with the rest of the parameters as the condition. The Gibbs sampler start with an initial guess  $\mathbf{X}^0 = [X_1^0, \dots, X_n^0]$  for the param-

eters, then iteratively update each as follows

$$\begin{aligned}
 X_1^{t+1} &\sim f(X_1|X_2^t, \dots, X_n^t), \\
 X_2^{t+1} &\sim f(X_2|X_1^{t+1}, X_3^t, \dots, X_n^t), \\
 &\vdots \\
 X_n^{t+1} &\sim f(X_n|X_1^{t+1}, \dots, X_{n-1}^{t+1}).
 \end{aligned} \tag{1.22}$$

where (t,X,f and n is....)

The iterative process is repeated until enough realizations are generated for sufficient accuracy.

REFFF!! (Gamerman and Lopes, 2006, p. 141)

(Givens and Hoeting, 2013, p. 209)

Geman and Geman (1984)

### 1.4.2 Metropolis–Hastings Algorithm

Metropolis–Hastings algorithm is another method for constructing a suitable Markov Chain. The algorithm is preferable for situations where a proportional distribution is simple to evaluate, while the target probability distribution is difficult. Bayesian inference (see chapter 1.3), often result in a distribution where the normalizing constant cannot analytically be calculated. While possible numerically, the normalizing constant often becomes computationally hard, which make it impracticable for iterative simulations. Using Metropolis–Hastings algorithm on a proportional distribution whiteout normalizing constant, results in samples from the target distributions.

The Metropolis–Hastings algorithm start with an initial guess for the parameters. For each iteration a new parameters  $\mathbf{X}^*$  are suggested from a proposal distribution  $g(\mathbf{X}^*|\mathbf{X}^t)$ , given the last accepted parameter  $\mathbf{X}^t$ . The new parameter is then evaluated against the last accepted by

$$R(\mathbf{X}^*, \mathbf{X}^t) = \frac{f(\mathbf{X}^*)g(\mathbf{X}^t|\mathbf{X}^*)}{f(\mathbf{X}^t)g(\mathbf{X}^*|\mathbf{X}^t)} \tag{1.23}$$

where  $f(x)$  is the target distribution, or a distribution proportional to the target distribution.

The parameter  $\mathbf{X}^{t+1}$  takes value  $\mathbf{X}^*$  with probability  $\min\{1, R(\mathbf{X}^*, \mathbf{X}^t)\}$ , if rejected,  $\mathbf{X}^{t+1} = \mathbf{X}^t$  instead. The reason the target distribution normalizing constant is irrelevant, is because they are both canceled out in  $f(\mathbf{X}^*)/f(\mathbf{X}^t)$ .

A common proposal distribution is the random walk. The new parameters are generated from the last accepted realization with additional variance,  $\mathbf{X}^* = \mathbf{X}^t + \boldsymbol{\epsilon}$  where  $\boldsymbol{\epsilon}$  follows a chosen probability distribution. Symmetric Proposals implies that  $g(\mathbf{X}^t|\mathbf{X}^*) = g(\mathbf{X}^*|\mathbf{X}^t)$ , this is referred to as Metropolis algorithm.

The Metropolis–Hastings algorithm could in situations be necessary for some of the steps in the Gibbs sampler, equation (1.22). Such a combination of Metropolis-Hastings algorithm and Gibbs sampler is referred to as a Hybrid Gibbs sampler.

(Givens and Hoeting, 2013, p. 202)

(Givens and Hoeting, 2013, p. 216)

(Gamerman and Lopes, 2006, p. 191)

(Gamerman and Lopes, 2006, p. 205)

Metropolis et al. (1953)

Müller (1991)

hybrid as in (1.22). REFFF!!

### 1.4.3 Effective Sample Size

The realizations of the simulated Markov Chain will often be correlated, and dependent on the future and past iterations. The correlation implies that the information hold by the realizations is actually less than the numbers of realizations. The effective sample size gives a method of calculating the theoretical size of an equally informative independent and identical distributed, realization set. The effective sample size is estimated as

$$L_{eff} = \frac{L}{1 + 2 \sum_{k=1}^K \hat{\rho}(k)}, \quad (1.24)$$

where  $L$  is the realizations of the sample size,  $\hat{\rho}(k)$  is the estimated  $k$  step autocorrelation between realizations and  $K$  is chosen as the first  $k$  where  $\hat{\rho}(k) < 0.1$  ref(BOK COMP STAT). (Coles,

2001, p. 174) The effective sample size is a quantification of the information hold by the realization set.

#### 1.4.4 Adaptive Metropolis Algorithm

A challenge with constructing a MCMC is to ensure that the series converge to the stationary target distribution relatively quickly, and that the samples gives points in the whole range of the target distribution, this is referred to as good mixing.

If a large percentage of Metropolis-Hastings proposal  $\mathbf{X}^*$  is accepted, the proposal distribution is too narrow. High acceptance rate will delay convergence, and cause higher correlation between points. The result is poor mixing and a decrease in effective sample size. On the other hand, if only a small percentage of the proposals is accepted, the proposal distribution is too wide. Low acceptance rate will also increase correlation, give poor mixing and decreased effective sample size. A large number of generated realizations by the Markov chain will be equal, which will harm future Monte Carlo simulation.

To maximize the effective sample size and ensure good mixing, the acceptance rate should be somewhere in between. For a Metropolis-Hastings algorithm it is suggested a 44% acceptance rate for single dimensional normal target distribution and 23.4% for high dimensional multivariate normal target distribution [Gelman et al. \(1996\)](#). Commonly the user would run the Metropolis-Hastings algorithm, calculate acceptance rate, tune variance and then rerun the process until sufficient acceptance rate is achieved.

For this paper/thesis, MCMC simulation will be used for a large number of different situations, and it would become extremely time-consuming to tune each variance. The inconvenience can be handled by using an Adaptive Markov Chain Monte Carlo (AMCMC) which adapts the MCMC algorithm while running. This is achievable using a normal random walk proposal where the next suggested realization  $\mathbf{X}^* \sim N(\mathbf{X}^t, \lambda \Sigma^t)$ . Between iterations  $\Sigma^t$  is adjusted to improve mixing and efficient sample size. The acceptance rate is set by  $\lambda$ , and with a  $p$  dimensional multivariate normal target distribution, it has been shown that a constant  $\lambda = 2.38^2/p$  is optimal when  $\Sigma$  equals the real variance of the target distribution, [Gelman et al. \(1996\)](#). The adaptive Metropolis algorithm is not constraint to the normal target distributions, but the suggested  $\lambda$  seems like a good starting value. The ability of an adjustable  $\lambda$  between future iteration

seems beneficial, because of the unknown target distribution and the following acceptance rate. The additional adaptive parameter  $\mu^t$  is necessary since the covariance is proportional to  $\mu$ . The initial guess is chosen as  $\mu^0 = \mathbf{0}$  and  $\Sigma^0 = \mathbf{I}$ . The normal random walk proposal distribution is symmetric, which result in an adaptive Metropolis algorithm, where (1.23) is reduced to

$$R(\mathbf{X}^*, \mathbf{X}^t) = \frac{f(\mathbf{X}^*)}{f(\mathbf{X}^t)}. \quad (1.25)$$

For each iteration  $\mu^{t+1}$  and  $\Sigma^{t+1}$  is updated as follows

$$\mu^{t+1} = \mu^t + \gamma^{t+1}(\mathbf{X}^{t+1} - \mu^t) \quad (1.26)$$

$$\Sigma^{t+1} = \Sigma^t + \gamma^{t+1}[(\mathbf{X}^{t+1} - \mu^t)(\mathbf{X}^{t+1} - \mu^t)^T - \Sigma^t], \quad (1.27)$$

Where  $\gamma$  is a decreasing parameters which provide the Markov chain property described in the beginning of chapter 1.4. The details of  $\gamma^t$ , to ensure an irreducible and aperiodic Markov chain can be read in ? and ? , where it is noted that  $\lim_{t \rightarrow \infty} \gamma^t = 0$  while the summation, not necessary is bounded  $\sum_{t=1}^{\infty} \gamma = \infty$ . Repeated trails concluded that  $\gamma^t = \dots$  was sufficient choice.

As described above an adaptive  $\lambda^t$  could be beneficial. By using

$$\log(\lambda^{t+1}) = \log(\lambda^t) + \gamma^{t+1}(R(\mathbf{X}^*, \mathbf{X}^t) - a), \quad (1.28)$$

the series acceptance rate will converge towards  $a$  (Givens and Hoeting, 2013, p. 248).

Gelman et al. (1996)

(Givens and Hoeting, 2013, p. 247)

### 1.4.5 Applying Markov Chain Monte Carlo to Peak Over Threshold

For the POT method.. (1.12) something something something SKRIV HER

The left side of equation ?? can be written as

$$\Pr(X > x | X > u) = \frac{\Pr(X > x, X > u)}{\Pr(X > u)}, \quad (1.29)$$

where the nominator  $\Pr(X > x, X > u) = \Pr(X > x)$ , since the conditional distribution require



$x > u$ . Reorganizing ?? gives

$$\Pr(X > x) = \Pr(X > u) \cdot \left[ 1 + \xi \left( \frac{x - u}{\sigma} \right) \right]^{-\frac{1}{\xi}}. \quad (1.30)$$

irreducible if its state space is a single communicating class; in other words, if it is possible to get to any state from any state. check.

aperiodic if there exists  $n$  such that for all  $n' > n$   $\Pr(X_{n'} = i | X_0 = i) > 0$  Usefull to have as much knowledge as possible about the target distribution as possible => better simulation ish 23% acceptance.

Prior distribution hard to choose.... future improvements!!

### 1.4.6 Multivariate Random Normal Generator

lkajsd yous tla blaa [Box and Muller \(1958\)](#)

## 1.5 Value at Risk

mer bla bla

## 1.6 ARCH/GARCH

is this needed????

# Chapter 2

## Data

hei

### 2.1 Synthetic Data

blabla

#### 2.1.1 Pareto

sim sim

#### 2.1.2 Commodity Generated Data

sam sam

### 2.2 Commodity Data

blabla

## **Chapter 3**

# **Summary and Recommendations for Further Work**

Better prior MCMC (Improved prior)

### **3.1 Summary and Conclusions**

Here, you present a brief summary of your work and list the main results you have got. You should give comments to each of the objectives in Chapter 1 and state whether or not you have met the objective. If you have not met the objective, you should explain why (e.g., data not available, too difficult).

This section is similar to the Summary and Conclusions in the beginning of your report, but more detailed—referring to the the various sections in the report.

### **3.2 Discussion**

Here, you may discuss your findings, their strengths and limitations.

### **3.3 Recommendations for Further Work**

You should give recommendations to possible extensions to your work. The recommendations should be as specific as possible, preferably with an objective and an indication of a possible approach.

The recommendations may be classified as:

- Short-term
- Medium-term
- Long-term

# Bibliography

- Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates. *The Annals of Mathematical Statistics*.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC, 2nd edition.
- Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). Efficient metropolis jumping rules. In *Bayesian Statistics 5*.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Givens, G. H. and Hoeting, J. A. (2013). *Computational Statistics*. John and Sons, 2nd edition.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*.
- Müller, P. (1991). A generic approach to posterior integration and gibbs sampling. Technical report, Department of Statistics, Purdue University.